Genomics and Its Impact on Medicine and Society

A 2001 Primer

U.S. Department of Energy Human Genome Program: www.ornl.gov/hgmis

The Basics

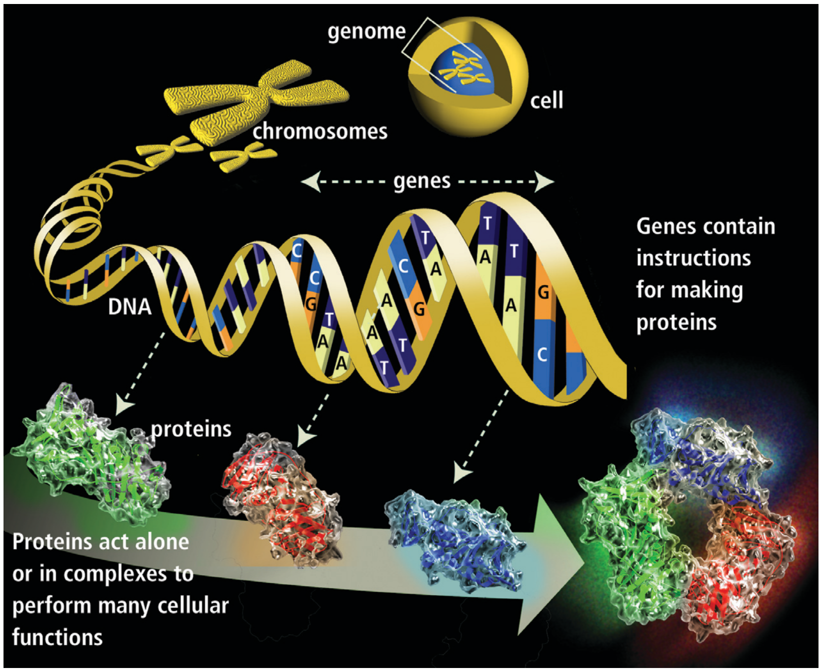
Cells are the fundamental working units of every living system. All the instructions needed to direct their activities are contained within the chemical DNA (deoxyribonucleic acid).

DNA from all organisms is made up of the same chemical and physical components. The DNA sequence is the particular side-by-side arrangement of bases along the DNA strand (e.g., ATTCCGGA). This order spells out the exact instructions required to create a particular organism with its own unique traits.

The **genome** is an organism's complete set of DNA. Genomes vary widely in size: the smallest known genome for a free-living organism (a bacterium) contains about 600,000 DNA base pairs, while human and mouse genomes have some 3 billion. Except for mature red blood cells, all human cells contain a complete genome.

DNA in the human genome is arranged into 24 distinct **chromosomes**—physically separate molecules that range in length from about 50 million to 250 million base pairs. A few types of major chromosomal abnormalities, including missing or extra copies or gross breaks and rejoinings (translocations), can be detected by microscopic examination. Most changes in DNA, however, are more subtle and require a closer analysis of the DNA molecule to find perhaps single-base differences.

Each chromosome contains many **genes**, the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. Genes comprise only about 2% of the human genome; the remainder consists of noncoding regions, whose functions may include providing chromosomal structural integrity and regulating where, when, and in what quantity proteins are made. The human genome is estimated to contain 30,000 to 40,000 genes.



Although genes get a lot of attention, it's the **proteins** that perform most life functions and even make up the majority of cellular structures. Proteins are large, complex molecules made up of smaller subunits called amino acids. Chemical properties that distinguish the 20 different amino acids cause the protein chains to fold up into specific three-dimensional structures that define their particular functions in the cell.

The constellation of all proteins in a cell is called its **proteome**. Unlike the relatively unchanging genome, the dynamic proteome changes from

moment to moment in response to tens of thousands of intra- and extracellular environmental signals. A protein's chemistry and behavior are specified by the gene sequence and by the number and identities of other proteins made in the same cell at the same time and with which it associates and reacts. Studies to explore protein structure and activities, known as proteomics, will be the focus of much research for decades to come and will help elucidate the molecular basis of health and disease.

The Human Genome Project—A Little Bit of History

Though surprising to many, the Human Genome Project (HGP) traces its roots to an initiative in the U.S. Department of Energy (DOE). Since 1945, DOE and its predecessor agencies have been charged by Congress to develop new energy resources and technologies and to pursue a deeper understanding of potential health and environmental risks posed by their production and use. Such studies have since provided the scientific basis for individual risk assessments, for example, of nuclear medicine technologies.

In 1986, DOE took a bold step in announcing its Human Genome Initiative, convinced that DOE's missions would be well served by a reference human genome sequence. Shortly thereafter, DOE and the National Institutes of Health developed a plan for a joint HGP that officially began in 1990.

Ambitious Goals . . .

From the outset, the HGP's ultimate goal has been to generate a high-quality reference sequence for the entire human genome and identify all human genes. Other important goals are to sequence the genomes of model organisms to help interpret human DNA, enhance computational resources to support future research and commercial applications, and explore gene function through mouse-human comparisons. Potential applications are numerous and include customized medicines, improved agriculture products, new energy resources, and tools for environmental cleanup.

The HGP also aims to train future scientists, study human variation, and address critical societal issues arising from the increased availability of personal human genome data and related analytical technologies.

. . . and Exciting Progress

Although the HGP originally was planned to last 15 years, rapid technological advances and worldwide participation have accelerated the expected completion date to 2003. In June 2000, scientists announced biology's most stunning achievement: the generation of a working draft sequence of the entire human genome. In addition to serving as a scaffold for the finished version, the draft provides a road map to an estimated 90% of genes on every chromosome and already has enabled gene hunters to pinpoint genes associated with more than 30 disorders.

HGP resources have spurred a boom in spin-off sequencing programs on the human and other genomes in both the private and public sectors. To stimulate further research, all data generated in the public sector are made available rapidly and free of charge via the Web.

HGP Spinoff Projects

Microbial Genome Project
www.sc.doe.gov/ober/microbial.html
www.ornl.gov/microbialgenomes

Microbial Cell Project
microbialcellproject.org

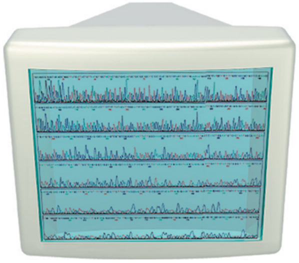
Genomes to Life
doegenomestolife.org

Environmental Genome Project
www.niehs.nih.gov/envgenom/home.htm

Cancer Genome Anatomy Project
www.ncbi.nlm.nih.gov/ncicgap

SNP Consortium
snp.cshl.org

A Major Milestone—Achieving the "Working Draft" Human Genome Sequence



In February 2001, HGP and Celera Genomics scientists published the long-awaited details of the working-draft DNA sequence. Although the draft is filled with mysteries, the first panoramic view of the human genetic landscape has revealed a wealth of information and some early surprises. Papers describing research observations in the journals *Nature* (Feb. 15, 2001) and *Science* (Feb. 16, 2001) are freely accessible (see www.ornl.gov/hgmis/project/journals/journals.html).

Although clearly not a Holy Grail or Rosetta Stone for deciphering all of biology—two early metaphors commonly used to describe the coveted prize—the sequence is a magnificent and unprecedented resource that will serve as a basis for research and discovery throughout this century and beyond. It will have diverse practical applications and a profound impact upon how we view ourselves and our place in the tapestry of life.

One insight already gleaned from the sequence is that, even on the molecular level, we are more than the sum of our 35,000 or so genes. Surprisingly, this newly estimated number of genes is only one-third as great as previously thought and only twice as many as those of a tiny transparent worm, although the numbers may be revised as more computational and experimental analyses are performed. At once humbled and intrigued by this finding, scientists suggest that the genetic key to human complexity lies not in the number of genes but in how gene parts are used to build different products in a process called alternative splicing. Other sources of added complexity are the thousands of chemical modifications made to proteins and the repertoire of regulatory mechanisms controlling these processes.

The draft encompasses 90% of the human genome's euchromatic portion, which contains the most genes. In constructing the working draft, the 16 genome sequencing centers produced over 22.1 billion bases of raw sequence data, comprising overlapping fragments totaling 3.9 billion bases and providing sevenfold coverage (sequenced seven times) of the human genome. Over 30% is high-quality, finished sequence, with eight- to tenfold coverage, 99.99% accuracy, and few gaps.

High-Quality Version Expected in 2003

The entire working draft will be finished to high quality by 2003. Coincidentally, that year also will be the 50th anniversary of Watson and Crick's publication of DNA structure that launched the era of molecular genetics (see www.nature.com/genomics/human/watson-crick). Much will remain to be deciphered even then. Some highlights follow from *Nature*, *Science*, and The Wellcome Trust philanthropy (an HGP funder).

Bioinformatics Boom: Managing the Data

Massive quantities of genomic data and high-throughput technologies are now enabling studies on a vastly larger scale than ever before, for example, in monitoring and comparing the activity of tens of thousands of genes simultaneously in cancerous and noncancerous tissue. Advanced computational tools and interdisciplinary experts are needed to capture, represent, store, integrate, distribute, and analyze all the data.

Bioinformatics is the term coined for the new field that merges biology, computer science, and information technology to manage and analyze the data, with the ultimate goal of understanding and modeling living systems. Computing and information demands will continue to rise with the explosive torrent of data from large-scale studies at the molecular, cellular, and whole organism levels (tour of the public DNA sequence database GenBank: www.ncbi.nlm.nih.gov/Tour/).

What Does the Working Draft Tell Us?

- The human genome contains 3164.7 million chemical nucleotide bases (A, C, T, and G).
- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
- The total number of genes is estimated at 30,000 to 40,000, much lower than previous estimates of 80,000 to 140,000 that had been based on extrapolations from gene-rich areas as opposed to a composite of gene-rich and gene-poor areas.
- The order of almost all (99.9%) nucleotide bases is exactly the same in all people.
- The functions are unknown for more than 50% of discovered genes.

The Wheat from the Chaff

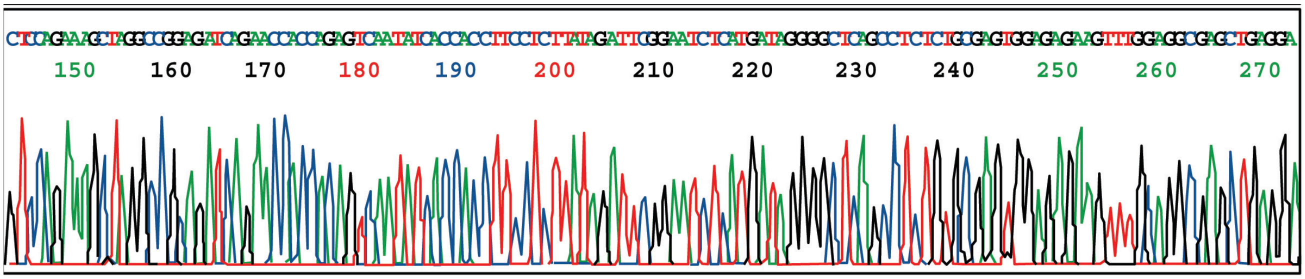
- About 2% of the genome encodes instructions for the synthesis of proteins.
- Repeated sequences that do not code for proteins ("junk DNA") make up at least 50% of the human genome.
- Repetitive sequences are thought to have no direct functions, but they shed light on chromosome structure and dynamics. Over time, these repeats reshape the genome by rearranging it, thereby creating entirely new genes or modifying and reshuffling existing genes.
- During the past 50 million years, a dramatic decrease seems to have occurred in the rate of accumulation of these repeats.

How It's Arranged

- The human genome's gene-dense "urban centers" are predominantly composed of the DNA building blocks G and C.
- In contrast, the gene-poor "deserts" are rich in the DNA building blocks A and T. GC- and AT-rich regions usually can be seen through a microscope as light and dark bands on the chromosomes.
- Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.
- Stretches of up to 30,000 C and G bases repeating over and over often occur adjacent to gene-rich areas, forming a barrier between the genes and the "junk DNA." These CpG islands are believed to help regulate gene activity.
- Chromosome 1 has the most genes (2968), and the Y chromosome has the fewest (231).

How the Human Genome Compares with Those of Other Organisms

- Unlike the human's seemingly random distribution of gene-rich areas, many other organisms' genomes are more uniform, with genes evenly spaced throughout.
- Humans have on average three times as many kinds of proteins as the fly or worm because of mRNA transcript "alternative splicing" and chemical modifications to the proteins. This process can yield different protein products from the same gene.



- Humans share most of the same protein families with worms, flies, and plants, but the number of gene family members has expanded in humans, especially in proteins involved in development and immunity.

- The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%).

- Although humans appear to have stopped accumulating repetitive DNA over 50 million years ago, there seems to be no such decline in rodents. This may account for some of the fundamental differences between hominids and rodents, although estimates of gene numbers are similar in both species. Scientists have proposed many theories to explain evolutionary contrasts between humans and other organisms, including those of life span, litter sizes, inbreeding, and genetic drift.

Variations and Mutations

- Scientists have identified about 1.4 million locations where single-base DNA differences (SNPs, see Sequence Variation, p. 11) occur in humans. This information promises to revolutionize the processes of finding chromosomal locations for disease-associated sequences and tracing human history.

More on the Working Draft

Papers from *Science* and *Nature*
www.ornl.gov/hgmis/project/journals/journals.html

Sequence Access Sites
www.ornl.gov/hgmis/project/journals/sequencesites.html

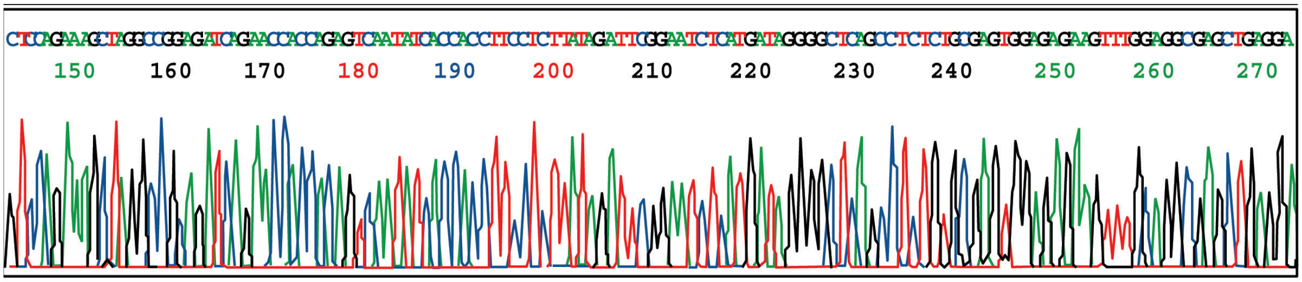
- The ratio of germline (sperm or egg cell) mutations is 2:1 in males vs females. Researchers point to several reasons for the higher mutation rate in the male germline, including the greater number of cell divisions required for sperm formation than for eggs.

Applications, Future Challenges

Deriving meaningful knowledge from the DNA sequence will define research through the coming decades to inform our understanding of biological systems. This enormous task will require the expertise and creativity of tens of thousands of scientists from varied disciplines in both the public and private sectors worldwide.

The draft sequence already is having an impact on finding genes associated with disease. Genes have been pinpointed and associated with numerous diseases and disorders including breast cancer, muscle disease, deafness, and blindness. Additionally, finding the DNA sequences underlying such common diseases as cardiovascular disease, diabetes, arthritis, and cancers is being aided by the human SNP maps generated in the HGP in cooperation with the private sector. These genes and SNPs provide focused targets for the development of effective new therapies.

One of the greatest impacts of having the sequence may well be in enabling an entirely new approach to biological research. In the past, researchers studied one or a few genes at a time. With whole-genome sequences and new automated, high-throughput technologies, they can approach questions systematically and on a grand scale. They can study all the genes in a genome, for example, or all the gene products in a particular tissue or organ or tumor, or how tens of thousands of genes and proteins work together in interconnected networks to orchestrate the chemistry of life.



After the HGP, the Next Steps . . .

The words of Winston Churchill, spoken in 1942 after 3 years of war, capture well the HGP era: "Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

The avalanche of genome data grows daily. The new challenge will be to use this vast reservoir of data to explore how DNA and proteins work with each other and the environment to create complex, dynamic living systems. Systematic studies of function on a grand scale—functional genomics—will be the focus of biological explorations in this century and beyond. These explorations will encompass studies in transcriptomics, proteomics, structural genomics, new experimental methodologies, and comparative genomics.

- **Transcriptomics** involves large-scale analysis of messenger RNAs (molecules that are transcribed from active genes) to determine when, where, and under what conditions genes are expressed.
- **Proteomics**—the study of protein expression and function—can bring researchers closer than gene-expression studies to what's actually happening in the cell.
- **Structural genomics** initiatives are being launched worldwide to generate the 3-D structures of one or more proteins from each protein family, thus offering clues to their function and providing biological targets for drug design.
- **Knockout studies** are one experimental method for understanding the function of DNA sequences and the proteins they encode. Researchers inactivate genes in living organisms and monitor any changes that could reveal the function of specific genes.
- **Comparative genomics**—analyzing DNA sequence patterns of humans and well-studied model organisms side by side—has become one of the most powerful strategies for identifying human genes and interpreting their function.

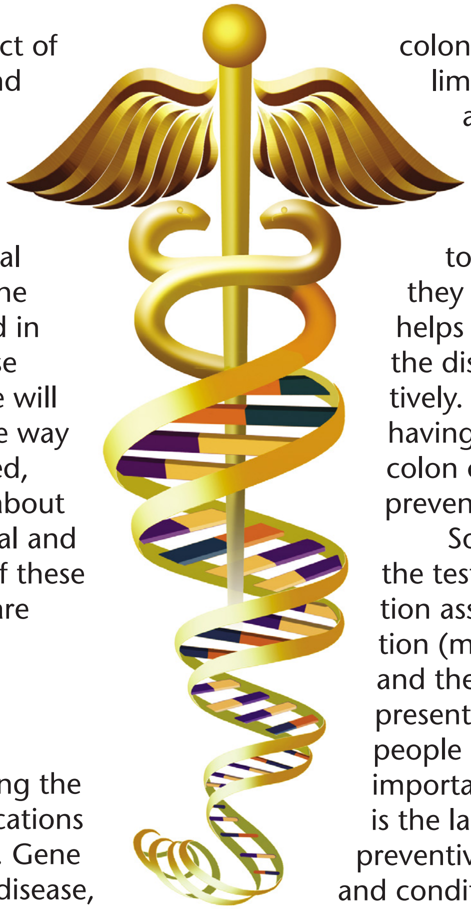
Today, scientists have in hand the complete DNA sequences of genomes for many organisms—from microbes to mice to humans. For the first time, we can begin to explore the "operating systems" of life written into these genetic codes and put them to use. At the leading edge of this great scientific frontier is the DOE Genomes to Life program, whose overarching goal is to use these biological tools to target critical mission challenges in energy independence, global climate change, toxic waste cleanup, and human health protection.

Future applications of this knowledge promise far-reaching benefits to the nation:

- Independence from foreign oil
- Significant savings in toxic waste cleanup and disposal
- Stabilization of atmospheric carbon dioxide to counter global warming
- Enhanced biowarfare agent detection and response

Genomes to life will build on the Human Genome Project, both by exploiting its data and by extending the whole-genome approach to biology to the next level—attaining a comprehensive understanding of how entire biological systems work. This will lead to models that enable scientists to understand living systems well enough to predict their behavior under different environmental conditions. Applications of this level of understanding will be revolutionary.

Medicine and the New Genetics: Gene Testing, Pharmacogenomics, Gene Therapy



DNA underlies every aspect of our health, both in function and dysfunction. Obtaining a detailed picture of how genes and other DNA sequences function together and interact with environmental factors ultimately will lead to the discovery of pathways involved in normal processes and in disease pathogenesis. Such knowledge will have a profound impact on the way disorders are diagnosed, treated, and prevented and will bring about revolutionary changes in clinical and public health practice. Some of these transformative developments are described below.

Gene Tests

DNA-based tests are among the first commercial medical applications of the new genetic discoveries. Gene tests can be used to diagnose disease, confirm a diagnosis, provide prognostic information about the course of disease, confirm the existence of a disease in asymptomatic individuals, and, with varying degrees of accuracy, predict the risk of future disease in healthy individuals or their progeny.

Currently, several hundred genetic tests are in clinical use, with many more under development, and their numbers and varieties are expected to increase rapidly over the next decade. Most current tests detect mutations associated with rare genetic disorders that follow Mendelian inheritance patterns. These include myotonic and Duchenne muscular dystrophies, cystic fibrosis, neurofibromatosis type 1, sickle cell anemia, and Huntington's disease.

Recently, tests have been developed to detect mutations for a handful of more complex conditions such as breast, ovarian, and colon cancers. Although they have limitations, these tests sometimes are used to make risk estimates in presymptomatic individuals with a family history of the disorder. One potential benefit to using these gene tests is that they could provide information that helps physicians and patients manage the disease or condition more effectively. Regular colonoscopies for those having mutations associated with colon cancer, for instance, could prevent thousands of deaths each year.

Some scientific limitations are that the tests may not detect every mutation associated with a particular condition (many are as yet undiscovered), and the ones they do detect may present different risks to different people and populations. Another important consideration in gene testing is the lack of effective treatments or preventive measures for many diseases and conditions now being diagnosed or predicted.

Revealing information about risk of future disease can have significant emotional and psychological effects as well. Moreover, the absence of privacy and antidiscrimination legal protections can lead to discrimination in employment or insurance or other misuse of personal genetic information. Additionally, because genetic tests reveal information about individuals and their families, test results can affect family dynamics. Results also can pose risks for population groups if they lead to group stigmatization.

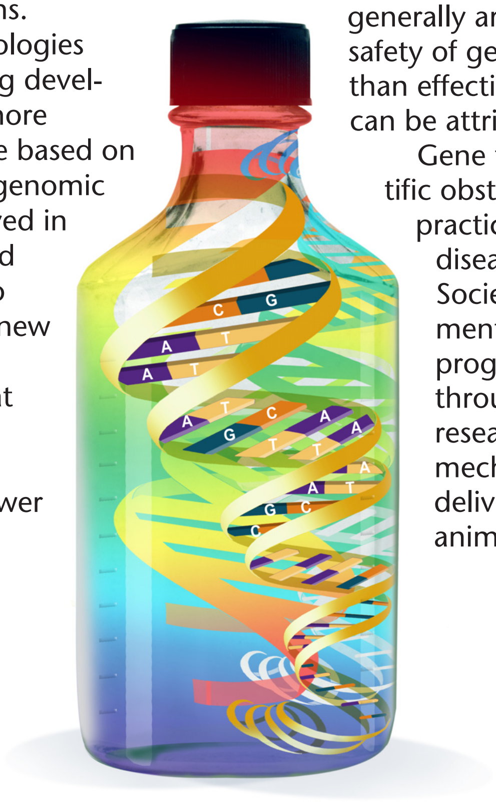
Other issues related to gene tests include their effective introduction into clinical practice, the regulation of laboratory quality assurance, the availability of testing for rare diseases, and the education of healthcare providers and patients about correct interpretation and attendant risks.

Pharmacogenomics: Moving Away from "One-Size-Fits-All" Therapeutics

Within the next decade, researchers will begin to correlate DNA variants with individual responses to medical treatments, identify particular subgroups of patients, and develop drugs customized for those populations. The discipline that blends pharmacology with genomic capabilities is called pharmacogenomics.

More than 100,000 people die each year as the result of adverse responses to medications that are beneficial to others. Another 2.2 million experience serious reactions, while others fail to respond at all. DNA variants in genes involved in drug metabolism, particularly the cytochrome P450 multigene family, are the focus of much current research in this area. Enzymes encoded by these genes are responsible for metabolizing most drugs used today, including many for treating psychiatric, neurological, and cardiovascular diseases. Enzyme function affects patients' responses to both the drug and the dose. Future advances will enable rapid testing to determine the patient's genotype and drastically reduce hospitalization resulting from adverse reactions.

Genomic data and technologies also are expected to make drug development faster, cheaper, and more effective. Most drugs today are based on about 500 molecular targets; genomic knowledge of the genes involved in diseases, disease pathways, and drug-response sites will lead to the discovery of thousands of new targets. New drugs, aimed at specific sites in the body and at particular biochemical events leading to disease, probably will cause fewer side effects than many current medicines. Ideally, the new genomic drugs could be given earlier in the disease process. As knowledge becomes available to select patients most likely to benefit from a potential drug, pharmacogenomics will speed the design of clinical trials to bring the drugs to market sooner.



Gene Therapy, Enhancement

The potential for using genes themselves to treat disease or enhance particular traits has captured the imagination of the public and the biomedical community. This largely experimental field—gene transfer or gene therapy—holds potential for treating or even curing such genetic and acquired diseases as cancers and AIDS by using normal genes to supplement or replace defective genes or bolster a normal function such as immunity.

More than 500 clinical gene-therapy trials involving about 3500 patients have been identified worldwide (June 2001). The vast majority (78%) take place in the United States, followed by Europe (18%). Most trials focus on various types of cancer, although monogenic, infectious, vascular, and other multigenic diseases are being studied as well. Protocols generally are aimed at establishing the safety of gene-delivery procedures rather than effectiveness, and no cures as yet can be attributed to these trials.

Gene transfer still faces many scientific obstacles before it can become a practical approach for treating disease. According to the American Society of Human Genetics' Statement on Gene Therapy, effective progress will be achieved only through continued rigorous research on the most fundamental mechanisms underlying gene delivery and gene expression in animals.

Other Anticipated Benefits of Genetic Research

Rapid progress in genome science and a glimpse into its potential applications have spurred observers to predict that biology will be the foremost science of the 21st century. Technology and resources generated by the Human Genome Project and other genomic research already are having major impacts on research across the life sciences. Doubling in size between 1993 and 1999, according to Ernst & Young,[1] the biotechnology industry generated 151,000 direct jobs and 287,000 indirect jobs. Revenues totaled $20 billion directly and $27 billion indirectly.

Some Current and Potential Applications of Genome Research

Molecular Medicine

- Improve diagnosis of disease
- Detect genetic predispositions to disease
- Create drugs based on molecular information
- Use gene therapy and control systems as drugs
- Design "custom drugs" based on individual genetic profiles

Microbial Genomics

- Rapidly detect and treat pathogens (disease-causing microbes) in clinical practice
- Develop new energy sources (biofuels)
- Monitor environments to detect pollutants
- Protect citizenry from biological and chemical warfare
- Clean up toxic waste safely and efficiently

Risk Assessment

- Evaluate the health risks faced by individuals who may be exposed to radiation (including low levels in industrial areas) and to cancer-causing chemicals and toxins

1. *The Economic Contributions of the Biotechnology Industry to the U.S. Economy*, May 2000, Ernst & Young.

Bioarchaeology, Anthropology, Evolution, and Human Migration

- Study evolution through germline mutations in lineages
- Study migration of different population groups based on maternal genetic inheritance
- Study mutations on the Y chromosome to trace lineage and migration of males
- Compare breakpoints in the evolution of mutations with ages of populations and historical events

DNA Identification

- Identify potential suspects whose DNA may match evidence left at crime scenes
- Exonerate persons wrongly accused of crimes
- Identify crime, catastrophe, and other victims
- Establish paternity and other family relationships
- Identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers)
- Detect bacteria and other organisms that may pollute air, water, soil, and food
- Match organ donors with recipients in transplant programs
- Determine pedigree for seed or livestock breeds
- Authenticate consumables such as caviar and wine

Agriculture, Livestock Breeding, and Bioprocessing

- Grow disease-, insect-, and drought-resistant crops
- Breed healthier, more productive, disease-resistant farm animals
- Grow more nutritious produce
- Develop biopesticides
- Incorporate edible vaccines into food products
- Develop new environmental cleanup uses for plants like tobacco

Societal Concerns Arising from the New Genetics

Since its inception, the Human Genome Project has dedicated funds toward studying the ethical, legal, and social issues (ELSI) surrounding the availability of the new data and capabilities. Examples of such issues follow.

- **Privacy and confidentiality of genetic information**. *Who owns and controls genetic information? Is genetic privacy different from medical privacy?*
- **Fairness in the use of genetic information** by insurers, employers, courts, schools, adoption agencies, and the military, among others. *Who should have access to personal genetic information, and how will it be used?*
- **Psychological impact, stigmatization, and discrimination** due to an individual's genetic differences. *How does personal genetic information affect self-identity and society's perceptions?*
- **Reproductive issues** including adequate and informed consent and the use of genetic information in reproductive decision making. *Do healthcare personnel properly counsel parents about risks and limitations? What are the larger societal issues raised by new reproductive technologies?*
- **Clinical issues** including the education—about capabilities, limitations, and social risks—of doctors and other health-service providers, people identified with genetic conditions, and the general public; and the implementation of standards and quality control measures. *How do we prepare health professionals for the new genetics? How do we prepare the public to make informed choices? How will genetic tests be evaluated and regulated for accuracy, reliability, and usefulness?* (Currently, there is little regulation at the federal level.) *How do we as a society balance current scientific limitations and social risk with long-term benefits?*
- **Fairness in access to advanced genomic technologies.** *Who will benefit? Will there be major worldwide inequities?*
- **Uncertainties associated with gene tests for susceptibilities and complex conditions** (e.g., heart disease, diabetes, and Alzheimer's disease). *Should testing be performed when no treatment is available or when interpretation is unsure? Should children be tested for susceptibility to adult-onset diseases?*
- **Conceptual and philosophical implications** regarding human responsibility, free will vs genetic determinism, and concepts of health and disease. *Do our genes influence our behavior, and can we control it? What is considered acceptable diversity? Where is the line drawn between medical treatment and enhancement?*
- **Health and environmental issues** concerning genetically modified (GM) foods and microbes. *Are GM foods and other products safe for humans and the environment? How will these technologies affect developing nations' dependence on industrialized nations?*
- **Commercialization of products** including property rights (patents, copyrights, and trade secrets) and accessibility of data and materials. *Will patenting DNA sequences limit their accessibility and development into useful products?*

Human Genome Project Goals 1998–2003

Human DNA Sequencing

The HGP's continued emphasis is on obtaining by 2003 a complete and highly accurate reference sequence (1 error in 10,000 bases) that is largely continuous across each human chromosome. Scientists believe that knowing this sequence is critically important for understanding human biology and for applications to other fields.

A "working draft" of the sequence was completed 18 months ahead of schedule, in June 2000. The achievement has provided scientists worldwide with a road map to an estimated 90% of genes on every chromosome. Although the draft contains gaps and errors and does not yet meet the criterion of 1 error in 10,000 bases outlined above, it provides a valuable scaffold for generating a high-quality reference genome sequence. HGP scientists make human DNA sequence available broadly, rapidly, and free of charge via the Web.

Sequencing Technology

Although current sequencing capacity is far greater than at the inception of the HGP, further incremental progress in sequencing technologies, efficiency, and cost-reduction are needed. For future sequencing applications, planners emphasize the importance of supporting novel technologies that may be 5 to 10 years in development.

Sequence Variation

Although more than 99% of human DNA sequences are the same across the population, variations in DNA sequence can have a major impact on how humans respond to disease; to such environmental insults as bacteria, viruses, toxins, and chemicals; and to drugs and other therapies.

Methods are being developed to detect different types of variation, particularly the most common type called single-nucleotide polymorphisms (SNPs), which occur about once every 100 to 300 bases. Scientists believe SNP maps will help them identify the multiple genes associated with such complex diseases as cancer, diabetes, vascular disease, and some forms of mental illness. These associations are difficult to establish with conventional gene-hunting methods because a single altered gene may make only a small contribution to disease risk.

Functional Genomics

Efficient interpretation of the functions of human genes and other DNA sequences requires that strategies be developed to enable large-scale investigations across whole genomes. A first priority is to generate complete sets of full-length cDNA clones and sequences for human and model-organism genes. Other functional-genomics goals include studies into gene expression and control and the development of experimental and computational methods for understanding gene function.

Comparative Genomics

The functions of human genes and other DNA regions often are revealed by studying their parallels in nonhumans. HGP researchers have obtained complete genomic sequences for the bacterium *Escherichia coli,* the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila melanogaster*, and the roundworm *Caenorhabditis elegans*. Sequencing continues on the laboratory mouse. The availability of complete genome sequences generated both inside and outside the HGP is driving a major breakthrough in fundamental biology as scientists compare entire genomes to gain new insights into evolutionary, biochemical, genetic, metabolic, and physiological pathways.

Ethical, Legal, and Social Implications (ELSI)

Rapid advances in the science of genetics and its applications present new and complex ethical and policy issues for individuals and society. ELSI programs that identify and address these implications have been an integral part of the U.S. HGP since its inception. These programs have resulted in a body of work that promotes education and helps guide the conduct of genetic research and the development of related medical and public policies.

Bioinformatics and Computational Biology

Continued investment in current and new databases and analytical tools is critical to the success of the HGP and to the future usefulness of the data it produces. Databases must adapt to the evolving needs of the scientific community and must allow queries to be answered easily. Planners suggest developing a human genome database, analogous to model organism databases, that will link to phenotypic information. Also needed are databases and analytical tools for studying the expanding body of gene-expression and functional data, for modeling complex biological networks and interactions, and for collecting and analyzing sequence-variation data.

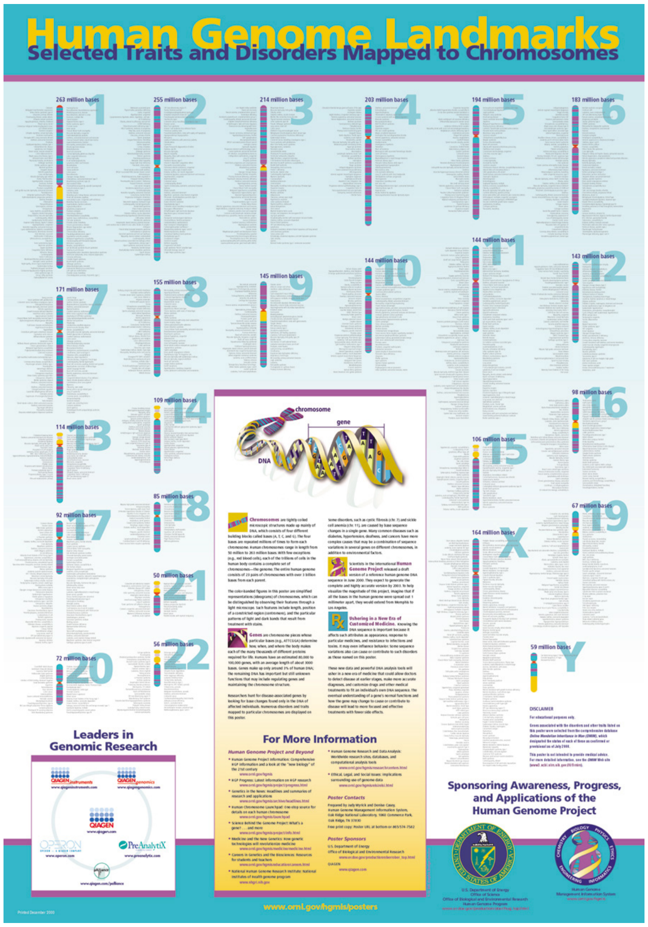
Training

Future genome scientists will require training in interdisciplinary areas including biology, computer science, engineering, mathematics, physics, and chemistry. Also, scientists with management skills will be needed for leading large data-production efforts.

For More Information

- **Human Genome Project Information**
  www.ornl.gov/hgmis
- **Online Version of this Primer**
  www.ornl.gov/hgmis/publicat/primer/intro.html
- **Medicine and the New Genetics**
  www.ornl.gov/hgmis/medicine/medicine.html
- **Ethical, Legal, and Social Issues**
  www.ornl.gov/hgmis/elsi/elsi.html
- **Genomes to Life (post-HGP research)**
  DOEGenomesToLife.org
- **Publication of Genome Draft Sequence**
  www.ornl.gov/hgmis/project/journals/journals.html
- **Image Gallery**
  www.ornl.gov/hgmis/education/images.html
- **Resources for Teachers**
  www.ornl.gov/hgmis/education/education.html
- **Resources for Students**
  www.ornl.gov/hgmis/education/students.html
- **Careers in Genomics**
  www.ornl.gov/hgmis/education/careers.html

Wall Poster Available Free of Charge



A wall poster depicting the 24 human chromosomes and many mapped genes may be ordered from www.ornl.gov/hgmis/posters/chromosome/ or from HGMIS. Informative sidebars explain genetic terms and provide URLs for finding more detailed information.

October 2001. Produced by the Human Genome Management Information System (HGMIS) at Oak Ridge National Laboratory, Oak Ridge, Tennessee, for the U.S. Department of Energy Human Genome Program.

For additional copies, contact: caseydk@ornl.gov, 865/574-0597, www.ornl.gov/hgmis.

A PDF version of this document and accompanying PowerPoint slides may be downloaded from www.ornl.gov/hgmis/publicat/primer/intro.html.